International Journal of Energy
ISSN: 2957-9473 | Vol. 2, No. 2, 2023

Research on oil and gas production prediction process based on machine learning

Zhenzhi Liu, Sanshan Li, Lu Li

Xi'an Shiyou University, Xi'an, Shaanxi 710065, China

Abstract: In recent years, the development trend of artificial intelligence is getting better and better. It has been widely used not only in the fields of big data analysis, automobile automatic driving, intelligent robot and face recognition, but also in various fields of oil and gas industry. Oil and gas production prediction is an important part of reservoir engineering, which is very important for the future production and development of strata, and can give developers some development suggestions. At present, the methods used in oil and gas production prediction are mainly traditional means such as numerical simulation and history matching. With the application of artificial intelligence in various fields of oil and gas industry, the use of machine learning models for oil and gas production prediction has become the direction of development and research. This paper summarizes the basic process and main technical means of applying machine learning model to predict oil and gas production by investigating the research of domestic and foreign scholars on artificial intelligence in oil and gas production prediction in recent years. It provides ideas and lays a foundation for future researchers to study this aspect, and also contributes to the development of smart oil fields in the future.

Keywords: Production forecast; Machine learning; Oil and gas; Artificial intelligence.

1. Introduction

With the advent and development of the information age, the combination of artificial intelligence and big data with various industries has become a development trend. Also, in the oil industry, the combination of big data and artificial intelligence has gradually become a trend [1-3]. For example, in oil and gas exploration, artificial intelligence machine learning model is used to analyze logging data, identify terrain structure and classify lithology [4-5]. In drilling, artificial intelligence algorithm is used to optimize drilling data, predict downhole data and combine with other drilling parameters to improve drilling efficiency [6-7]. In terms of oil and gas development, artificial intelligence is used to predict the formation fracture pressure and analyze the main controlling factors of oil and gas production [8-9]. In the field of reservoir engineering, machine learning algorithms are used to predict oil and water distribution [10] and oil and gas production. In other areas of oil and gas such as the economy, there are also a wide range of applications, such as dynamic forecasting of oil prices [11].

At present, most of the methods used to predict oil and gas production are traditional methods, such as history matching, numerical simulation and decline curve analysis. These traditional methods not only consume energy, but also have some disadvantages. The first is the acquisition of data. The traditional method has high requirements for data. When using the traditional method to predict the yield, all the parameters required by the method need to be obtained first, and the data processing method is insufficient. The data cannot be analyzed in detail and the parameters can be screened to a certain extent. Especially for the well site with very complex geological conditions, the data are numerous and messy. If the data processing is not done well, the accuracy of the model prediction will be greatly affected. The second problem is that after a certain period of production, the data changes greatly and the output fluctuates greatly. The traditional method cannot obtain new data and new trend information for fitting prediction. The most important problem is that the traditional prediction method cannot analyze the prediction results, cannot analyze the influencing factors of production, and cannot give the follow-up development opinions of production developers through data analysis.

In order to solve or optimize these problems, researchers began to use machine learning in artificial intelligence to predict oil and gas production. Compared with traditional methods, artificial intelligence has great advantages. First, there are many data processing methods that can greatly optimize the collected data, and some parameters that cannot be learned by the model can also be redefined as new parameters. Secondly, the model learning ability is strong, the analysis ability of nonlinear data is strong, and the good model can accurately learn and predict the trend of data transformation. There are also many analysis methods. After the results are obtained, a series of visual processing can be performed on the results, the influencing factors can be analyzed, and the developer 's subsequent development opinions can be given to optimize the cost savings.

2. The Basic Process and Technical Means of Oil and Gas Production Prediction Based on Machine Learning

At present, using machine learning to predict oil and gas production is mainly two aspects. One is to make static prediction, that is, using geological features or engineering features as input of the model to predict the cumulative production of oil and gas or some indicators that can characterize oil and gas production[12-18]; the second is to carry out dynamic prediction, that is, not only using geological features or engineering features as the input of the model, but also setting the time step, taking the historical yield trend with time series as the feature, to carry out dynamic yield simulation[19-28]. Through the investigation

of the above two applications, the basic process and technical means of oil and gas production prediction based on machine learning are summarized.

2.1. Data Collection

Data collection is the basis for modeling, so accurate data is crucial for model training and prediction. One of the methods of collecting data is to collect the data of the actual oil and gas field, which is more in line with the actual production and can better reflect the role of machine learning in oil and gas prediction. Another method of data collection is to numerically simulate the geology and production of oil and gas fields, and obtain data through simulation. Such data errors are relatively small, and the model can be better trained and predicted.

2.2. Data Processing

After the data is collected, the data needs to be processed, so that the data can be more accurate and complete, and the model can be predicted quickly and accurately. There are several kinds of data processing:

1)Data cleaning

The actual collected data is often vacant or error due to human reasons, and the collection of some data is difficult. Therefore, the data is cleaned up, such as the processing of vacancy values and noise data, so as to obtain data more suitable for the model. Vacancy data processing is to fill the vacant data or directly abandon the vacant data. Noise processing is to smooth the data and reduce the influence of noise data on model prediction.

2)Normalization

For some models, the data need to be normalized, which can improve the convergence speed of the model. The two most commonly used methods are as follows

i. linear normalization, which converts the data into the range of [0,1]. The formula is:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

where the normalized data set is X_{norm}, the original data set is X, the minimum value of the data set is $X_{\min}$, and the maximum value of the data set is $X_{\max}$.

ii. zero mean standardization, the original data set is normalized to a data set with a mean of 0 and a variance of 1. The formula is:

$$Z = \frac{x - \mu}{\delta}$$

where the original data set is μ, the method is δ.

3)Variable correlation processing

We can visually view the correlation between features through visualization. For variables with large correlation, one or several variables can be discarded. If we are particularly familiar with the data set and understand the physical relationship between the parameters, we can perform manual screening to analyze the relationship between variables and reduce the number of parameters. We can also judge the correlation between variables by Pearson correlation coefficient. The range of the Pearson coefficient is between 0 and 1. When the Pearson coefficient is 0, there is no correlation between the two variables. When the Pearson coefficient is closer to 1, the correlation between the two variables is stronger.

4)Other processing

the number of parameters can be reduced according to the physical properties or formulas between data to improve the speed of the model. For example: the length and width of the rectangle, we can combine the two parameters into the perimeter or area of the rectangle.

After processing the data, the data set is divided into training set and test set. Sometimes a validation set will be separated.

2.3. Model Choice

After data processing, we can choose the model according to the actual situation. At present, there are many kinds of commonly used models. After a lot of literature research, the following are summarized:

Linear Regression, Support Vector Machine, Random Gradient Descent, K-nearest Neighbor, Ridge Regression, Extreme Random Forest, Gradient Boosting Tree, Random Forest, CatBoost, LGBM, Decision Tree and Neural Network Algorithm. The Neural Network Algorithm includes a variety of algorithms, such as BP Neural Network, Convolutional Neural Network, Long Short-term Memory Network and so on.

Due to the uncertainty of data and the diversity of models, it is impossible to accurately determine which model the predicted data is suitable for. Therefore, multiple models should be trained at the same time when predicting, and then the predicted results should be compared to select the most suitable model.

2.4. Model Training

After the model is determined, we can use the divided training data set to train the model. During training, a crucial process is to optimize the hyperparameters of the model. The three main hyperparameter optimizations include the following four:

1)Manual parameter adjustment

Manual parameter adjustment is to input the parameters of the model directly. For people with a lot of model training experience, this method is a very accurate and fast tuning method. But for the experience is not enough or novice, this method is not practical, not only will waste a lot of energy, and even lead to the failure of training.

2)Grid search

Search all the values within the specified range to determine the optimal parameters. Because of its simplicity, this method is currently the most widely used method. If the search step is small and the search range is large, this method can easily locate the optimal hyperparameters, but it also consumes a lot of time and energy.

3)Random search

This method is essentially the same as grid search, which is to search within the scope. However, the difference is that random search does not search all the values in the search range as grid search, but randomly selects them. This method has a certain probability to select the optimal parameters in a wide range of search, but the results of this algorithm are uncertain. Usually, in the process of hyperparameter optimization, it is used as a fast version of grid search. Firstly, it is used to search, determine the approximate hyperparameter position, and then determine a certain range near its value, and then perform grid search. In this way, the combination of the two methods can accelerate the speed of searching hyperparameters.

4)Bayesian optimization

This method is different from random search and grid search, which makes full use of the information searched previously. He can draw on the previous results to influence the subsequent optimization. It will first search according to the prior distribution, and then a new search, will use the search results more.

2.5. Model Prediction

After the model is trained, it can be predicted according to the divided test set.

2.6. Result Analysis

After the prediction, we will compare and analyze the results of the selected models. When analyzing the results, we need to evaluate the quality of the model through a variety of evaluation indicators. After a lot of literature research, summed up the evaluation index of the regression model we often use:

Mean Absolute Error (MAE)

$$MAE = \frac{1}{\mathrm{n}} \sum_{t=1}^{n} \left| y(t) - y(t) \right|$$

where the actual value is $y(t)$, the predicted value is $y(t)$.

Mean SquareError (MSE)

$$MSE = \frac{1}{\mathrm{n}} \sum_{t=1}^{\mathrm{n}} \left(y(t) - y(t) \right)^2$$

where the actual value is $y(t)$, the predicted value is $y(t)$.

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{MSE}$$

The root mean square error is the square root of the mean square error.

Coefficient of Determination (R2)

$$R^2 = SSR / SST = 1 - SSE / SST$$

where the sum of total squares is SST, the sum of regression squares is SSR, and the sum of residual squares is SSE.

2.7. Other Analysis

After evaluating the model, we can also conduct other visual analysis of the optimal model and data based on the existing artificial intelligence technology, so as to find out the characteristics that have the greatest impact on the model, that is, the greatest factor affecting oil and gas production. It can also analyze how the selected features affect the model or production, so as to give developers some opinions.

1)Feature importance

This is a way to score the features (independent variables) of the model input, and then show the contribution of each feature (independent variable) to the model when the model is trained and predicted. Feature importance analysis can make us better understand our data sets. Through the analysis of feature importance, we can clearly determine which feature contributes more to the prediction of the model, which contributes less to the model, or even close to zero. For features that are close to zero or contribute too little compared to other features, we can choose to discard such features to optimize the model, speed up the training of the model, and improve performance. There are three main sources for obtaining feature importance:

2)Characteristic importance coefficient

First, a model is fitted, and then the coeff-attribute of the model is output. This attribute contains the coefficients found for each feature, and these coefficients can roughly determine the importance of the feature.

3)Feature importance based on decision tree model

In various models based on decision tree, such as random forest, extreme random forest, etc., there is an attribute feature _importance, which contains the importance information of all features.

4)Random sorting feature importance

This method is suitable for models that do not support the importance of output features, so it is independent of the model used. We can use the permutation _importance () function to randomly sort the output of feature importance.

5)Interpretability analysis

Using this analysis method, the influence relationship between the input parameters of the model and the prediction of the model can be analyzed, which provides a basis for the improvement of the model and parameter optimization. Common interpretability analysis methods are SHapley Additive exPlanations, Partial Dependence Plot and Individual Conditional Expectation Plot.

6)Sample size analysis

Sample size analysis can draw a line chart of the impact of sample size on the model prediction results, so as to determine whether the sample size we selected when training the model is the best amount, and the trend of sample size on the model.

3. Conclusion

After a lot of literature research and analysis, the basic process and technical means of oil and gas production prediction based on machine learning are summarized. At the same time, it can be seen from the process that compared with the traditional prediction method, machine learning in artificial intelligence has better advantages and practical value in oil and gas production prediction.

Firstly, the modeling speed of oil and gas production prediction model based on machine learning is faster than that of traditional methods, and has higher accuracy.

Secondly, the oil and gas production prediction model based on machine learning has a variety of data processing methods, and has a fast processing speed, which can greatly reduce the noise of the data and improve the training speed of the model.

Finally, in the analysis of the results, machine learning provides a lot of visual analysis, which not only allows us to better understand the prediction of the model, but also provides development suggestions for developers, and even improves oil and gas production.

In summary, it is feasible and efficient to use machine learning to predict oil and gas production.

References

[1] MIN C, DAI B R, ZHANG X H, et al. A Review of the Application Progress of Machine Learning in Oil and Gas Industry[J]. Journal of Southwest Petroleum University(Science & Technology Edition), 2020, 42(6): 1- 15.

[2] Kuang L C, Liu H, Ren Y L, et al. Application and development trend of artificial intelligence in petroleum exploration and development[J]. Petroleum Exploration and Development. 2021, 48(1): 1- 11.

[3] Cheng Y F,Yang Y. Application of Artificial Intelligence in Oil Well Production Prediction[J]. Chemical Engineering Design Communications, 2021, 47(01): 125- 126.

[4] Guo Y. Feature recognition from potential fields using neural networks[J]. SEG Technical Program Expanded Abstracts, 1949, 11(1): 1410.

[5] Gupta K D, Vallega V, Maniar H, et al. A deep-learning approach for borehole image interpretation[C]//SPWLA 60th Annual Logging Symposium. OnePetro, 2019.

[6] Ashena R, Rabiei M, Rasouli V, et al. Drilling parameters optimization using an innovative artificial intelligence model[J]. Journal of Energy Resources Technology, 2021, 143(5).

[7] Aliyev R, Paul D. A novel application of artificial neural networks to predict rate of penetration[C]//SPE Western Regional Meeting. OnePetro, 2019.

[8] Kiss A, Fruhwirth R K, Pongratz R, et al. Formation breakdown pressure prediction with artificial neural networks[C]//SPE International Hydraulic Fracturing Technology Conference and Exhibition. OnePetro, 2018.

[9] Lin X, Liu Z S, Gao Y, et al. Analysis of main control factors of oil production based on machine learning [J]. China CIO News, 2019.

[10] Gu J W, Ren Y L, Wang Y K, et al. Prediction methods of remaining oil plane distribution based on machine learning[J]. Journal of China University of Petroleum (Edition of Natural Science), 2020, 44(4): 39- 46.

[11] Xu Z J. Forecastion oil price trends with news sentiment[D]. Beijing University of Chemical Technology, 2016.

[12] Li T, Tan Y, Ahmad F A, et al. A new method to production prediction for the shale gas reservoir[J]. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2020: 1-14.

[13] Lin B, Guo J, Liu X, et al. Prediction of flowback ratio and production in Sichuan shale gas reservoirs and their relationships with stimulated reservoir volume[J]. Journal of Petroleum Science and Engineering, 2020, 184: 106529.

[14] Meng M, Zhong R, Wei Z. Prediction of methane adsorption in shale: Classical models and machine learning based models[J]. Fuel, 2020, 278: 118358.

[15] Amirian E, Dejam M, Chen Z. Performance forecasting for polymer flooding in heavy oil reservoirs[J]. Fuel, 2018, 216: 83- 100.

[16] A Amirian E, Dejam M, Chen Z. Performance forecasting for polymer flooding in heavy oil reservoirs[J]. Fuel, 2018, 216: 83-100.

[17] Amirian E, Fedutenko E, Yang C, et al. Artificial neural network modeling and forecasting of oil reservoir performance[J]. Applications of Data Management and Analysis: Case Studies in Social Networks and Beyond, 2018: 43-67.

[18] Li J H, Ji L. Productivity forecast for multi-stage fracturing in shale gas wells based on a random forest algorithm[J]. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2020: 1-10.

[19] Nguyen-Le V, Shin H, Little E. Development of shale gas prediction models for long-term production and economics based on early production data in barnett reservoir[J]. Energies, 2020, 13(2): 424.

[20] Xue L, Liu Y, Xiong Y, et al. A data-driven shale gas production forecasting method based on the multi-objective random forest regression[J]. Journal of Petroleum Science and Engineering, 2021, 196: 107801.

[21] Wang H L, Mu L X,Shi F G, et al.Production prediction at ultra-high water cut stage via Recurrent Neural Network[J]. Petroleum Exploration and Development, 2020, 47(05): 1084-1090.

[22] Zhang R, Jia H. Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for waterflooding reservoirs[J]. Petroleum Exploration and Development, 2021, 48(01): 175-184.

[23] Liu W, Liu W, Gu J W. Oil production prediction based on a machine learning method[J]. Oil Drilling & Production Technology, 2020, 42(01): 70- 75.

[24] Sagheer A, Kotb M. Time series forecasting of petroleum production using deep LSTM recurrent networks[J]. Neurocomputing, 2019, 323: 203- 213.

[25] Gu J W, Zhou M, Li Z T, et al. Oil Well Production Forecast with Long-Short Term Memory Network Model Based on Data Mining[J]. Special Oil & Gas Reservoirs, 2019, 26(02): 77-81+ 131.

[26] Hou C H. New well oil production forecast method based on long-term and short-term memory neural network[J]. Petroleum Geology and Recovery Efficiency, 2019, 26(03): 105- 110.

[27] Hua J C. Production forecast modeling and application based on machine learning [D]. China University of Geosciences (Beijing), 2020.

[28] Lu Z Y. Research on production forecast method of gas well in tight gas reservoir based on big data analysis[D]. Southwest Petroleum University, 2019.